# Limitations on testing quantum theory

**Fang Zhang**[1], **Cupjin Huang**[1], **Michael Newman**[2], **Kevin Sung**[1] **& Yaoyun Shi**[1,3]

[1]Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48103, USA

[2]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

[3]Alibaba Group, Bellevue, WA, USA

fangzh@umich.edu, cupjinh@umich.edu, mgnewman@umich.edu
kevjsung@umich.edu, y.shi@alibaba-inc.com

**Abstract**

How much of quantum theory can be experimentally tested? We investigate this question in an extreme scenario, where the experimenter cannot make any assumptions on the quantum description of a device or state. Instead, she can only directly manipulate and interpret classical information. We prove that in this scenario, there is almost nothing that she can ascertain about the internal state of a given quantum system. More precisely, for any initial state of a multi-partite system, there exists a different internal state so that any experiment performed on the system will produce identical statistics under these different states. In particular, we cannot certify if two subsystems are entangled. Our result implies that asserting the identity of a quantum state requires assumptions that cannot be experimentally tested through classical information alone.

## Introduction

- Self-testing techniques allow one to certify the existence of entanglement *between* untrusted devices [2].
- Can we certify entanglement between untrusted *states* too?
  - Both store quantum information, but...
  - Untrusted devices can generate classical outputs on its own;
  - Untrusted states must be input into an untrusted device if we want retrieve any information.
- If we simply input the states to the devices, we cannot determine whether the entanglement is between the input states or pre-existing between the devices themselves.
- Actually, there is no way to certify "anything interesting" about untrusted states.
  - Adversarial implementations can make it so that *all untrusted quantum states are in a maximally mixed joint state*[1].

## Attacks

- In this work, an "attack" is defined as adversarial untrusted devices "simulating" other "honest" devices with *identical classic behavior* in a protocol.
- This work describes two "attacks":
  - A "one-shot attack", which only ensures that all quantum inputs to the protocol are in the maximally mixed state.
  - An extended attack, which also makes sure that all untrusted quantum states *at any time point* are in the maximally mixed state.

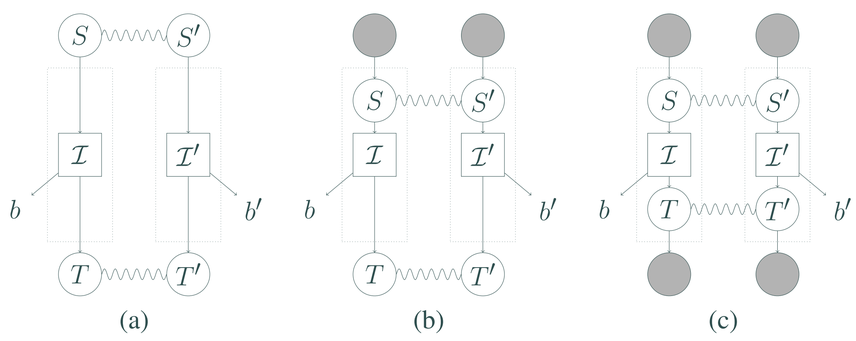


**Figure 1:** A high-level depiction of the simulation scheme. (a) The ideal case. For simplicity, the internal states of the devices are not represented here, and the "ideal box" $\mathcal{I}$ implements input ($S$ and $S'$), classical output ($b$ and $b'$) and quantum output ($T$ and $T'$) at once. (b) The one-shot attack. The input states become maximally mixed states, and $S$ and $S'$ are only recovered "inside the box". (c) The extended attack. Now the output states $T$ and $T'$ are also only found "inside the box" before being encoded into maximally mixed states. No matter which scenario we are in, the distributions of the classical outputs $b$ and $b'$ do not change.

## The One-Shot Attack

- Basic idea: encode all quantum inputs with a quantum secret sharing scheme [1].
  - Take an $((n, 2n-1))$ threshold scheme (where $n$ is the number of untrusted devices), put $n-1$ shares into the untrusted state, and allocate one share to each untrusted device.
- Any one untrusted device plus the untrusted device can recover the original state.
- All untrusted devices together can recover the original state even without the untrusted state.

## Extension of the Attack

- Encode the outputs from the untrusted devices as well as the original input.
- The devices cannot communicate directly with each other, but they can teleport some states nonetheless.
  - Some classical information is necessary to complete the teleportations...
  - Just bundle them with the untrusted state! They looks just like random strings!

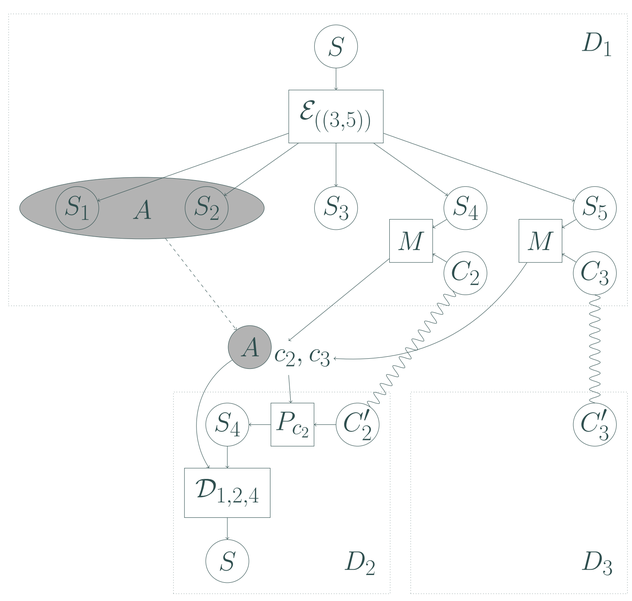


**Figure 2:** An illustration of how the devices can use teleportation to encode the quantum outputs. Here, $S$ is a quantum state that is supposed to be output by $D_1$, but instead $D_1$ encodes $S$ with a secret sharing scheme, and begins teleporting $S_4$ and $S_5$ by doing Bell measurements, denoted by $M$ in the diagram. Strictly speaking, $D_1$ should also teleport $S_3$ to itself, but of course this is unnecessary. The results of the measurements, $c_2$ and $c_3$, are included in the output as a classical tag (we omit the index of the entangled pairs used). When the output is given to $D_2$, it can perform the necessary Pauli correction according to $c_2$ and recover $S_4$, and then recover $S$ from $A$ and $S_4$. Notice that $c_3$ is never used and $S_5$ will never be recovered, but this is okay, as one quantum state cannot be input into two devices anyway.

## An Alternative Quantum Secret Sharing Scheme

- There is a more qubit-efficient way to implement the quantum secret sharing scheme.
- Basic idea: apply a quantum one-time pad to a qubit with a *coherent key*.
  - Interesting fact: the information in the "data" qubit is *transferred* to the "key" qubits, and can be extracted without using the "data" qubit.

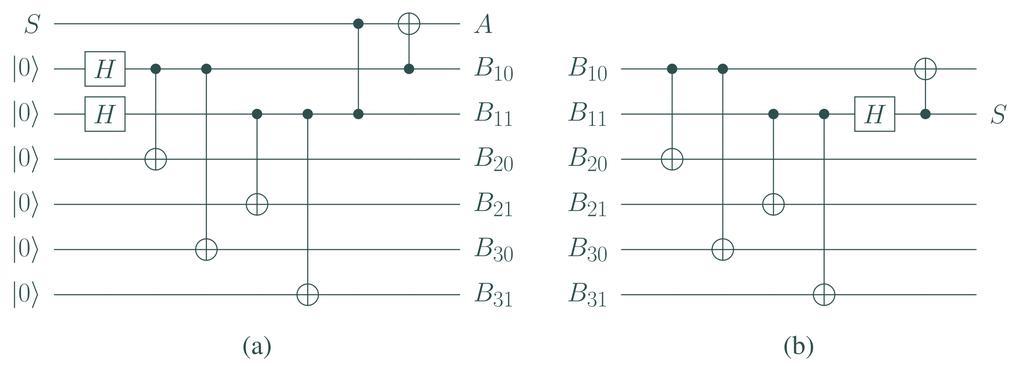


**Figure 3:** The secret sharing scheme based on the quantum one-time pad. (a) The encoding circuit for $n = 3$. The first two Hadamard gates and four CNOT gates generates two instances of the GHZ state, $(|000\rangle + |111\rangle)/\sqrt{2}$; the last two gates coherently apply the quantum one-time pad to $S$. Notice that in this circuit, the first four CNOT gates are interchangeable with the last two gates. Each share $B_i$ consists of two qubits. It is straightforward to recover $S$ given $A$ and any one $B_i$. (b) The circuit to recover $S$ from $\{B_1, B_2, B_3\}$. The first four CNOT gates just reverse the action of the first four CNOT gates in (a); the last two gates are the "core circuit". These circuits can easily be extended to the case of arbitrary $n$.

## Conclusions

- In theory, there are multiple possible "interpretations" of the world that are all consistent with quantum mechanics, but disagree on the quantum states.
- For device-independent protocols, one has to be content with classical outputs unless other assumptions are added (such as trusted input states, or trusted but noisy devices).
  - The classical outputs (which we can read) are not affected: this attack is so "perfect" that it will do the correct thing even with the wrong quantum states!

## References

[1] Richard Cleve, Daniel Gottesman, and Hoi-Kwong Lo. How to share a quantum secret. *Physical Review Letters*, 83(3):648, 1999.

[2] Carl A Miller and Yaoyun Shi. Optimal robust quantum self-testing by binary nonlocal xor games. *arXiv preprint arXiv:1207.1819*, 2012.

[1]Strictly speaking, in our model, untrusted quantum states may contain a classical "tag", which will be the only part that is not in a maximally mixed state. Of course, since the tags are classical, they cannot contribute any entanglement, so the main result "we cannot certify entanglement" is still true.